Are basic emotions ingrained in the brain's architecture?

Conrado Eiroa Solans, Department of Experimental Psychology

Abstract – Emotions are a series of functional states that promote specific action tendencies. Classical theorists postulated that individual emotions are ontologically distinct, and that as such, they are associated with specific sets of mental, behavioural, and physiological manifestations. However, inconsistencies in these phenomenological patterns have led some researchers to question whether instead, the distinctness lies on the broad dimensions of affect such that underpin emotions. Does the neurological evidence suggest that emotions are fundamentally different? In this essay, I take stock of neurological research on emotions and evaluate the extent to which current theory can help us answer this question. To do so, I first outline the fundaments of the classical and social-constructionist views of emotion, then explore a lesion case study and review neuroimaging evidence, and finally integrate these findings to shed light on the conceptual and empirical complexities underpinning the nature of emotions. Ultimately, I conclude that there is enough evidence to suggest that some aspects of emotion are fundamentally distinct.

Keywords: emotion, basic emotions, social constructionism, lesion, neuroimaging

Two Perspectives

Every day, we, as humans, automatically combine stimuli, life experience, and billions of years of evolution to satisfy our biological and psychological needs. During infancy, reflexes guide reactions to basic stimuli that human evolution has determined beneficial or harmful, and during childhood and adolescence, ambiguous or neutral stimuli become associated with positive or negative outcomes (e.g. Skinner, 1984). Over time, we develop increasingly nuanced responses and gain the ability to act volitionally, yet we never decide how to behave in a vacuum (Bullock & Lütkenhaus, 1988; Izard, 2007). We perceive the world in relation to our

needs, and from the moment we are born, functional states create evolutionarily adapted *tendencies* that help us to navigate complex environments in need of fast, self-relevant decisions (Frijda, 1986).

The classical view of emotion posited that these tendencies are basic emotions; discrete, functional affective states that can be differentiated through sets of recognisable, observable properties (Adolphs, 2017a). According to this perspective, emotions are biologically driven affective states (Frijda, 1986) that provide intermediate flexibility for complex behaviour (Adolphs, 2017a). Unlike reflexes, which create a direct connection between input and output, and volitional behavior, in which motor output is independent from sensory input, emotions encourage evolutionarily adapted action tendencies that influence cognition and predispose certain behavioural outputs (Izard, 1992; Adolphs & Anderson, 2018). Thus, classical theorists propose that emotions are tendencies that allow us to deal with environmental challenges in an efficient and context-sensitive manner.

Basic emotions could have evolved from reflexes as action tendencies due to their utility in addressing recurring problems faced by our ancestors (Frijda, 1986; Tooby & Cosmides, 1990). Indeed, they motivate and regulate both cognition and action in order to achieve an adaptive advantage (e.g. Ekman & Davidson, 1994; Öhman & Mineka, 2001). For instance, Panksepp (1998) found that all mammals share seven innate 'primary emotional systems' that predispose action related to competence, reproduction or self-preservation. The neurobiological systems responsible for such affective expression appear early in development (Buck, 1999), and many of these systems become functional during the first few years of age (Camras et al., 1998). Together, the existence and early development of a neurobiological system that fosters emotional reactions before higher cognition matures is a strong indicator that we have a basic set of evolutionarily determined affective responses.

However, mounting evidence of overlapping phenomenological properties has led some researchers to hypothesise that this might not be the case. Specifically, the behavioural characteristics that were previously thought to universally represent individual emotions seemed to be

impacted by cultural differences (Gendron et al., 2014); scales that were built to measure different emotions showed high correlations between similar basic emotions (e.g. anger and fear; Barrett, 2006); the autonomic manifestations of basic emotions reveal no distinctive patterns[1] (Kreibig, 2010); and initial meta-analyses of neurological bases of basic emotions proved to be inconsistent (e.g. Murphy et al., 2003 vs. Phan et al., 2002). Thus, the inability to differentiate emotions on the basis of mental, behavioural, and physiological properties led some researchers (e.g. Barrett, 2006) to depart from the classical view of emotion and question whether there are any bases to believe that individual emotions are fundamentally different.

The theory of socially constructed emotion postulates that our brains and bodies evolved to experience broad dimensions of affect, and that emotions per se are better defined as the internal interpretation of these dimensions (Barrett, 2006). As an analogy, we might think of emotions as colours: our eyes and brain work in tandem to sense varying wavelengths of light, but we interpret and communicate our appraisal of a particular wavelength by categorising colours on the basis of our previous experience and cultural preferences (Davidoff, 2001). Similarly, Barrett (2006) proposes that we can interoceptively sense varying degrees of an affective predisposition to certain events, but we classify these experiences into discrete emotions based on past experience, cognitive schemas, and culture. This view can explain cultural differences in emotion (e.g. Gendron et al., 2014), and why scales built to measure basic emotions showed high correlations between similar emotions (cf. Barrett, 2006); being exposed to similar categories of events throughout our lifespan and experiencing similar affective responses to those events may promote the impression that all humans experience discrete categories of similar emotions. However, according to this theory, emotions per se do not have evolutionarily pre-determined neural circuits responsible for producing

[1] For instance, anger, disgust and fear (three of the six basic emotions) share similar changes in most psychophysiological measures (i.e. heart rate, heart rate variability, systolic and diastolic blood pressure, electrodermal activity, respiratory rate, and even the adrenergic receptors they activate). Aggravatingly, happiness and anticipatory sadness (theoretically very different emotions) share more changes in common (heart rate, heart rate variability, respiration rate, systolic and diastolic blood pressure, skin conductance level, alpha adrenergic receptors) than anticipatory and acute sadness (different states of the same emotion; only respiratory rate; cf. Kreibig et al., 2010).

specific emotional experiences. Instead, social constructionists propose that these circuits belong to broader dimensions of affect, and emotions are more akin to cognitive concepts such as money (Barrett, 2017).

Indeed, it is also possible that nature selected for a more elementary affective system in the brain. Approaching what benefits us, and shunning away from what harms us, are the guiding principles of all human behaviour (Schneirla, 1959), so rather than discrete emotional states, evolution could have promoted the development of *broad affective systems* that use sensory systems and mental schemas to inform the organism whether a stimulus is harmful (i.e. good or bad; valence) and to prepare the organism for action if an immediate response is required (i.e. arousal; cf. Russell, 1980; cf. Chikazoe et al., 2014). According to this view, the behavioural direction of the action could be informed by both *behavioural inhibitory systems* (BIS), responsible for sensitivity to punishment and avoidance motivation, and *behavioural approach systems* (BAS), responsible for sensitivity to reward and approach motivation (Gray, 1981). In line with this idea, Barrett (2006) proposes that the brain represents broad dimensions of valence, arousal, and BIS/BAS generally referred to as core affect – not individual emotions.

Specifically, Barrett (2017) proposes that self-referential systems transform core affect into individual emotions. To accomplish this, the default mode network, a series of regions associated with self-referential processing and autobiographical memory retrieval and imagining the future (Ekhtiari et al., 2016) first forms an interoceptive sensation in the form of valence and arousal. Then, the salience system, a network of regions that perceives and responds to homeostatic demands (Seeley, 2019) matches our current goals with relevant sensory information about the world. Lastly, the frontoparietal control network, a system involved in sustained attention, problem solving, and working memory (Menon, 2015) creates a simulation of the expected sensations as well as the best action plan to maintain physical and psychological equilibrium. Together, these systems create individual mental states that drive specific action plans, and with experience, Barrett argues, we learn to categorise these mental states into distinct concepts we perceive as emotions.

The fundamental difference between these two perspectives is that classical theorists see emotions as *adaptive functional states* that arise in response to specific events, whilst constructionists argue that emotions are *predictive concepts* that prospectively regulate behaviour in the service of allostasis. The former explore whether there are discrete affective states that drive biological tendencies; the latter posit that the biological differences are limited to broad dimensions of affect and provide a framework explaining how emotion-concepts emerge.

Yet, there is no conclusive evidence on the ontological nature of emotions. The data suggest that emotions cannot be reliably dissociated on the basis of self-reported or behavioural measures, but studies investigating the peripheral profiles of individual emotions (e.g. heart rate variability, skin conductance level) have not found consistent evidence for or against either position. For instance, happiness and anticipatory sadness, two emotions that live at opposite sides of the valence spectrum, share more peripheral correlates than anticipatory and acute sadness (see Footnote 1; cf. Kreibig et al., 2010), which are merely two different forms of the same emotion. This evidence violates the classical assumption that basic emotions have distinct physiological manifestations. However, it is also incongruent with constructionist claims that the biological bases of emotion are grounded on basic forms of affect, since oppositely valenced emotions appear to have more in common in terms of peripheral correlates than two forms of an emotion with identical valence, arousal, and VIS/VAS profile.

Thus, to explore whether there are any biological grounds to believe that emotions are ontologically distinct, we must resort to extant evidence on the last frontier of biology: the human brain. Due to its role in giving birth to emotions and its structural complexity, the brain is uniquely poised to answer this question. If emotions have clearly dissociable structural and functional neurological substrates, then there would be biological grounds to claim that emotions are ontologically distinct – and vice versa. In the next sections, I discuss a lesion case study and review recent neuroimaging evidence to shed some light on the conceptual and empirical complexities underpinning this foundational debate.

A case study

Patient S.M. is a woman born with a particularly severe case of Urbach-Wiethe disease, an extremely rare genetic condition in which hyaline deposits build up on skin, mucosa, and internal organs (Hofer, 1973). One of the mysterious effects of this mutation is the selective calcification and consequent bilateral destruction of the amygdala – a lesion that has prevented S.M. from experiencing fear but not any other emotion. The severity of her lesion – alongside her willingness to collaborate with researchers – has enabled the scientific community to expand our understanding of the amygdala's role in experiencing fear, and exemplified the vital role individual neural structures can play in emotion (Feinstein et al., 2016).

S.M.'s condition developed throughout her childhood. She claims to have autobiographical memories of being afraid as a child, before the disease had fully affected her amygdala; but now, as an adult, the only way she can experience an emotional state similar to fear is by breathing in air with higher concentrations of carbon dioxide (Feinstein et al., 2011, 2013). S.M. has accumulated a wealth of conceptual knowledge of fear through books, movies, and conversations to the point where she can recognise fear in bodies and voices, correctly use the concept of fear in conversation, and tell you about the causes and manifestations of experiencing fear (Feinstein et al., 2016); yet despite exhibiting normal skin conductance responses to unexpectedly loud noises, her brain does not use arousal as a learning cue in fear-learning paradigms, nor show loss aversion when gambling, nor indeed show any of the normal physiological manifestations of fear (Feinstein et al., 2011, 2016; Barrett, 2017). S.M. appears to have successfully developed a *concept* of fear, but her ability to experience emotion-state fear as an adult is, at least, extremely limited.

There are, however, some facets of fear that S.M. might be able to experience without the amygdala. For instance, after a traumatically painful visit to the dentist, S.M. experienced anticipatory anxiety and strong feelings of worry at the thought of experiencing similar pain upon returning to repeat the procedure (Feinstein et al., 2016). Additionally, she asked the experimenters not to repeat the CO2 experiment because she

had experienced a very unpleasant flashback to the time her ex-husband choked her after she tried to confront him about an infidelity (Feinstein et al., 2016). In both situations, she denied feeling scared both during the events (instead experiencing pain or anger, respectively) and at the prospect of repeating the procedures. However, the fact that she exhibited avoidance behaviour to highly arousing, negatively valenced events shows that she could in fact experience key properties of emotion-state fear, even if she did not associate those properties with her concept of experiencing fear.

This raises an important question about the relationship between emotion-states and our subjective emotional experiences. S.M. appears to be capable of experiencing key aspects of the emotion-state society has agreed to name 'fear' (i.e. avoidance behaviour to highly arousing, negatively valenced events), but she denies *feeling* scared despite having a well-constructed emotion-concept of fear. Thus, could S.M. be, in fact, misidentifying the emotion-state she was experiencing? The neuroscientific literature does not support this view. Her neural lesion extends to regions that have reciprocal relationships with the amygdala such as the ventromedial prefrontal cortex, which is strongly related to the processing of risk and fear (Motzkin et al., 2015), and the enthorhinal cortext, a region that is associated with the ability to condition temporally separated stimuli (Ryou et al., 2001; cf. Feinstein et al., 2016; cf. Barrett, 2017). These functions lie at the core of the ability to experience what we know as 'fear', so it seems unlikely that without them S.M. can experience emotion-state fear.

Broadly, evidence from S.M. reveals three important facts:[2]

1. First, state-fear is distinct from other emotional states. It appears that the complete calcification of S.M.'s amygdala prevented her from experiencing fear, but it did not impact her ability to experience any other emotions.

[2] In a related example, there is evidence that lesions to the insula and basal ganglia prevent people from selectively experiencing disgust, but not any other emotion (Calder et al., 2000). Similar evidence on other emotions is yet to arise, but together, evidence on the role of the amygdala in fear, and the insula and basal ganglia in disgust, reveal in a promising pattern.

2. Second, state-fear is distinct from the concept of fear. S.M. had no trouble *understanding* the concept of fear, identifying it in others, or even mentalising it, but her experience with state-fear is anecdotally limited to memories from her early life.

3. Finally, fear can be divided into different components. S.M. seems to experience fear-related concepts and feelings, but even if she is in a highly aroused, negatively valenced state, she does not consciously experience fear.

Neuroimaging Debate

Interpreting neuroimaging data can be tricky. Unlike lesion studies, which allow us to draw causal inferences about whether neural structures are necessary for emotions, neuroimaging studies attempt to find the areas that are related to emotion from fully correlational data while using a technology that is vulnerable to sporadic correlations (Bennett et al., 2009; LeDeux, 2012). Thus, when assessing neuroimaging studies, it is particularly important to have a clear understanding of how the data could support a hypothesis. Initially, theorists hypothesised that emotions lay within specific regions of the limbic system (e.g. MacLean, 1952), but subsequent neuroimaging evidence revealed that no specific brain region is uniquely activated for each emotion (c.f. Murphy et al., 2003; Vytal & Hamann, 2010). Some theorists claimed that disproving the lack of one-to-one correspondence between brain regions and emotions nullified one of the main characteristics of basic emotions (Barrett, 2017; cf. Ekman, 1999) and that as such, emotions were not likely to be ontologically different. However, one-to-one correspondence is not the only way emotions could manifest neurological uniqueness. Classical and constructionist theorists agree that the word 'emotion' should be used in the same way we use the words 'vision' or 'memory' (Barrett, 2017; Adolphs, 2017a). Yet there is no strict one-to-one relationship between brain regions and vision (Hubel & Wiesel, 1979; Pitzalis et al., 2018) or memory (e.g. Anderson, 1972), and the neurological uniqueness between vision and memory have yet to be questioned. Thus, lack of one-to-one correspondence between emotions and brain regions alone should not automatically preclude individual

emotions from being distinct (cf. Anderson, 2016). Instead, to identify whether basic affective operations in the brain are best represented by basic emotions or broad affective dimensions, we must look beyond the phrenological study of emotion and seek whether there are unique *patterns* in neural activation that represent emotional states.

Multi-voxel pattern analysis (MVPA) techniques have made great progress in finding such patterns (Vytal & Hamann, 2010; Saarimäki et al., 2016; Celeghin et al., 2017; Kirby & Robinson, 2017). By comparing data from more than one voxel at a time, MVPA has the power to analyse broader patterns of activity and can achieve a higher sensitivity than conventional univariate analysis (Mahmoudi et al., 2012). A significant problem in the study of the neural bases of emotion is that the methodologies used to induce emotions were responsible for a significant amount of variability across studies (Barrett & Wager, 2006), but studies using MVPA techniques have found unique neural signatures for each emotion that are independent from how the emotions were induced (Saarimäki et al., 2016; Celeghin et al., 2017; Bush et al., 2018). Through the lens of MVPA, the initially surprising findings that the amygdala is not only related to fear but also to anger, disgust, sadness, and joy (Sergerie et al., 2008; Lindquist et al., 2012; Tettemanti et al., 2012; Kirby & Robinson, 2017) can be explained by the idea that emotions reside not within specific regions, but within specific *interactions* between regions (Pessoa, 2014). For instance, amygdala activity has been found to correlate with posterior visual areas during the processing of fear and with the dorsomedial prefrontal cortex during the processing of happiness (Diano et al., 2017). The idea of emotions residing in networks of regions as opposed to specific regions in the brain, has opened up the door to new possibilities, and is helping us make sense of previously incongruent findings.

However, Clark-Polner and colleagues (2016; including Barrett) interpret this evidence differently. They argue that biological categories involve highly variable individuals, and statistical patterns that distinguish one category member from another do not necessarily exist in nature. They take the lack of one-to-one consistency between categories and their members to mean that emotions are highly variable across individuals, and while categorising emotions in networks can provide a valid statistical

summary, the authors counterargue that it does not prove that emotions are naturally distinct. In fact, they argue that this view aligns best with Barrett's theory of constructed emotion. Barrett hypothesised that the medial prefrontal and posterior cingulate cortices worked alongside limbic regions to produce allostatic predictions that we interoceptively perceive as emotions (see the Embodied Predictive Interoception Coding model; Barrett, 2012; Barrett & Simmons, 2015); Saarimäki et al.'s (2016) findings reinforce Barrett's hypotheses by showing that there are cortical and subcortical interactions during the experience of emotion, and that the medial prefrontal and posterior cingulate cortices are two key regions for the subjective experiences of most emotional states. Thus, instead of interpreting Saarimäki et al.'s results as evidence that basic emotions lie within specific interactions, constructionists specifically argue that these data reflect a *combination* of neural networks involved in affective interoception and emotion-concepts, rather than specific networks that represent basic emotions (Clark-Polner et al., 2016).

Nevertheless, there is a significant limitation to Barrett's arguments. The claim that statistical patterns do not reflect natural entities supports the idea that the classic view of emotion may have fallen into essentialism, but it neglects alternative explanations. Most notably, it is possible that the lack of one-to-one consistency observed in neuroimaging studies is related to individual differences in the emotional concepts that are automatically activated alongside emotions (Panksepp, 1994; Damasio, 1999; Izard, 2007). While constructionists group together emotions and emotion-concepts, classical theorists point out that in the same way 'concepts of planets are not planets, concepts of emotions are not emotions' (Adolphs, 2017b); emotions and emotion-concepts are fundamentally different things. Emotion-concepts are theorised to be primarily distributed across cognitive or perceptual areas, the regions that show most inconsistencies between emotion categories (Izard, 2007; cf. Saarimäki et al., 2015; cf. Kirby & Robinson, 2017), whereas emotion-states are thought to lie within networks of subcortical regions and possibly driven by a key subcortical structure (e.g. the amygdala for S.M.'s experience of fear; cf. Feinstein et al., 2016). Hence, while the inclusion of emotion-concepts in the definition of emotion might create a lens through which emotions do not have

identifiable neural underpinnings, there are theoretical and empirical grounds on which to separate the two, and not enough evidence to show that the neurological networks associated with emotion-states are not distinct.

In fact, Saarimäki and colleagues (2016) did find that unique networks were associated with individual emotions, and a recent study combining MVPA with magnetoencephalography (MEG; a nascent technology with greater temporal resolution than fMRI) showed that individual emotions also exhibit unique temporal patterns (see Grootswagers et al., 2021). Thus, while the activation of the medial prefrontal and posterior cingulate cortices with every emotion may reflect the interoceptive systems hypothesised by Barrett (Barrett, 2012; cf. Clark-Polner et al., 2016), it seems that this network is merely assisting with the interoception of individual emotions. In other words, Saarimäki and colleagues' (2016) data does support Barrett's (2012) hypothesis, but it *also* provides evidence that individual emotions are associated with unique networks of regions. Future research using MVPA and MEG technologies should continue to disentangle the seemingly overlapping patterns of activity in the limbic system and ultimately dissociate the neural substrates of emotion-states from emotion-concepts. For now, however, it would be premature to preclude the existence of individual emotion-states that exist independently from emotion-concepts (Izard, 2007; c.f. Wager et al., 2015).

Discussion

Emotion is a rapidly evolving concept. Evidence from lesion studies reveals that certain brain regions are fundamental for experiencing certain emotions, but neuroimaging technology shows that the neural representations of emotions are more variable than previously envisioned. Although this understanding has outdated Ekman's (1999) definition of basic emotions, theorists have struggled to encompass extant evidence on the nature of emotion in a single framework. Contrary to the classical view of emotion, it seems that basic emotions can be dissociated into more elementary facets of those emotions (see the case of S.M. section above; cf. Feinstein et al., 2016; cf. Adolphs, 2017a), and contrary to social

constructionist views, there is enough evidence to suggest that emotions are mental states that are neurologically different from each other, and that – as a whole – these functional states are qualitatively different from mental concepts that can be compared to 'money' (Barrett, 2017).

Lesion studies provide strong evidence for the classical view of emotion. S.M. is a clear example of 1) the extent to which fear concepts are different from state-fear; 2) the fact that fear is distinct from other emotions; and 3) the fact that fear can be separated into different sub-components. If we considered emotions as concepts, we could argue that the fear-related concepts S.M. has acquired throughout the years constitute an emotional experience, but electrophysiological data from fear-learning paradigms, behavioural data in loss-aversion tasks, and self-report data clearly indicate that she does not experience many of the markers that differentiate fear from other emotions (Feinstein et al., 2016). To be sure, the fact that she can experience negative valence, high arousal, and avoidance behaviours (i.e. at the prospect of going to the dentist, or breathing air rich in carbon dioxide) indicates she can still experience some of its physiological aspects, but the fact that in those scenarios S.M. reports feeling pain or anger due to the neurological damage she has experienced speaks to the importance of the amygdala and complimentary regions (e.g. ventromedial prefrontal cortex and enthorhinal cortex) in producing *state* fear. Thus, because a lesion to the amygdala prevents S.M. from experiencing state-fear, but not other aspects of fear (e.g. fear concepts) or indeed other emotions, it is likely that the neural substrates of fear are biologically unique.

Other non-neuroimaging approaches also support the importance of specific neural regions to emotion. S.M.'s case study is unique in that she suffers from a particularly severe case of a rare disease, but the data from this study are only part of the broader literature showing the importance of specific brain regions in relation to emotions. For instance, researchers have found that lesions to the insula and basal ganglia prevent people from selectively experiencing disgust, but not any other emotion (Calder et al., 2001); specific neural populations can produce specific-emotion behaviour (Lin et al., 2011); fear does not rely on cognitive processes for its activation (Öhman & Mineka, 2001; Mineka & Öhman, 2002); and direct brain stimulation can elicit discrete emotions (e.g. Krack et al., 2001; Okun et al.,

2004). Together, these non-neuroimaging findings suggest that there is indeed a strong relationship between brain regions and individual emotions, meaning that the neurological substrates of individual emotions are sufficiently unique that – in theory – if we mapped all of these regions and stimulated the appropriate areas, we should be able to artificially induce the functional state that underpins each emotional experience.

Neuroimaging evidence presents a slightly different picture. Based on current data, Barrett (2017) builds a compelling case that limbic regions govern elementary functions related to arousal, saliency, and relevance, and cortical regions help shape these emotional states into emotions (Barrett, 2017). However, her argument falters when she interprets this as evidence that emotions do not have specific neurological underpinnings. Neuroimaging evidence does suggest that becoming cognitively aware of our emotion-states involves interoceptive processes that make use of systems such as the default mode, salience, and frontoparietal control networks (Damasio & Carvalho, 2013; Kragel & LaBar, 2014; Barrett, 2017), but there is little evidence indicating that emotion-states do not have distinguishable neural systems beneath the neural representations of these emotion-concepts. In fact, there is some evidence from MVPA studies suggesting that emotion-states lie *within* specific interactions in the brain (Saarimäki et al., 2016; Celeghin et al., 2017; Bush et al., 2018). Additionally, there is nascent evidence from an MEG study using MVPA that different emotions exhibit unique temporal dynamics (Grootswagers et al., 2021), which suggests that the nature of these interactions may not only be physical, but also temporal. As things stand, constructionists have interpreted the current lack of one-to-one consistency found in neuroimaging studies as proof that significant findings are but statistical summaries, and that emotions are not to be found in nature. However, given most interpersonal differences are found in cortical regions that are also associated with emotion-concepts, and that there is evidence linking emotion-states to unique physical and temporal representations in the brain, it is possible that emotion-states (but not emotion-concepts) consistently rely on unique patterns of neurological activity.

Together, lesion and neuroimaging evidence suggest that core affect is probably a reducible reality in the brain. Lesion studies have shown that

damage to the amygdala can prevent people from selectively experiencing state-fear (Feinstein et al., 2016) and damage to the insula can impair people's ability to experience state-disgust (Calder et al., 2000). Neuroimaging research shows that these regions are vital to the salience (Ekhtiari et al., 2016) and default mode (Seeley, 2019) networks, respectively, which – as previously explored – play a key role in interocepting emotions from core affect (Barrett, 2017). If broad affective systems were the fundamental reality of the brain, and emotions were best described as mentalisations of such core affect, a lesion to one of these systems should damage at least whole clusters of similar emotions – but this is not what the data show. Despite fear and disgust being extremely similar by all measures of core affect (i.e. both are negatively valenced, high arousal, BIS-driven emotions), lesions to the amygdala exclusively impact people's ability to experience state-fear and lesions to the insula exclusively impact people's ability to experience state-disgust (cf. Calder et al., 2000; cf. Feinstein et al., 2016). Thus, given that there are key regions without which people can experience all emotion-states but one, and there is not such known relationship between brain regions and emotion-concepts, it seems likely that emotion-states are affective experiences qualitatively different from emotion-concepts that merit individual distinction.

Conclusion

Early theorists postulated that emotions are evolutionarily adapted affective states that promote specific action tendencies. Inconsistent evidence concerning their neurological underpinnings prompted some theorists to hypothesise that perhaps emotions were better construed as mentalisations of core affective states. Based on neuroimaging data, Barrett (2006, 2017) makes a compelling argument that subcortical regions might govern elementary aspects of emotion such as arousal, saliency, and relevance while cortical regions help to shape these emotional states into the emotion-concepts we experience. However, recent neuroimaging findings indicate that emotions are underpinned by unique physical and temporal patterns of region-activation. Additionally, lesion evidence indicates that certain subcortical brain regions are vital to the proper

experiencing of emotion-states. Thus, it is possible that individual regions are responsible for different *aspects* of the emotion, and that basic emotion-states are found within unique but overlapping networks of regions. In this manner, our emotional experiences could arise from a unique collection of properties, with emotion-states eliciting consistently distinct combinations of neural networks, and emotion-concepts activating highly variable representations in cortical regions that allow us to create, update and implement allostatic schemas. Future research should aim to delineate the exact relationship between emotion-concepts and emotion-states and continue to explore the unique dynamics underpinning individual emotion-states, but as things currently stand, there is enough evidence to suggest that emotion-states are grounded in the brain's architecture.

References

Adolphs, R., & Anderson, D. J. (2018). *The neuroscience of emotion: A new synthesis*. Princeton University Press.

Adolphs, R. (2017a). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social cognitive and affective neuroscience*, 12(1), 24–31.

Adolphs, R. (2017b). Reply to Barrett: affective neuroscience needs objective criteria for emotions. *Social cognitive and affective neuroscience*, 12(1), 32–33.

Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical biosciences*, 14(3-4), 197–220.

Anderson, M. L. (2016). Précis of after phrenology: neural reuse and the interactive brain. *Behavioral and Brain Sciences*, 39.

Barrett, L. F. (2006). Are emotions natural kinds?. *Perspectives on psychological science*, 1(1), 28–58.

Barrett, L. F. (2012). Emotions are real. *Emotion*, 12(3), 413–429.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1), 1–23.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature reviews neuroscience*, 16(7), 419–429.

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, 15(2), 79–83.

Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Neuroimage*, 47 (Suppl 1), S125.

Buck, R. (1999). The biological affects: a typology. *Psychological review*, 106(2), 301.

Bullock, M., & Lütkenhaus, P. (1988). The development of volitional behavior in the toddler years. *Child development*, 664–674.

Bush, K. A., Privratsky, A., Gardner, J., Zielinski, M. J., & Kilts, C. D. (2018). Common functional brain states encode both perceived emotion and the psychophysiological response to affective stimuli. *Scientific reports*, 8(1), 1–10.

Calder, A. J., Lawrence, A. D., & Young, A. W. (2001). Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, 2(5), 352–363.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience*, 3(11), 1077–1078.

Camras, L. A., Oster, H., Campos, J., Campos, R., Ujiie, T., Miyake, K., ... & Meng, Z. (1998). Production of emotional facial expressions in European American, Japanese, and Chinese infants. *Developmental Psychology*, 34(4), 616.

Celeghin, A., Diano, M., Bagnis, A., Viola, M., & Tamietto, M. (2017). Basic emotions in human neuroscience: neuroimaging and beyond. *Frontiers in Psychology*, 8, 1432.

Chikazoe, J., Lee, D.H., Kriegeskorte, N., & Anderson, A.K. (2014). Population coding of affect across stimuli, modalities, and individuals. *Nature Neuroscience*, 17(8), 1114–1122.

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, 27(3), 1944–1948.

Damasio, A.R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness.* New York: Harcourt Brace.

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature reviews neuroscience*, 14(2), 143–152.

Davidoff, J (2001). Language and perceptual categorization. *Trends in Cognitive Sciences*, 5(9), 382–387.

Diano, M., Tamietto, M., Celeghin, A., Weiskrantz, L., Tatu, M. K., Bagnis, A., ... & Costa, T. (2017). Dynamic changes in amygdala psychophysiological connectivity reveal distinct neural networks for facial expressions of basic emotions. *Scientific reports*, 7(1), 1–13

Ekman, P. (1999). Basic emotions. In T. Dalgleish and M. J. Power (Eds.). *Handbook of Cognition and Emotion* (pp. 45–60), New York, NY: John Wiley & Sons Ltd.

Ekhtiari, H., Nasseri, P., Yavari, F., Mokri, A., & Monterosso, J. (2016). Neuroscience of drug craving for addiction medicine: From circuits to therapies. *Progress in brain research*, 223, 115–141.

Ekman, P. E., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions*. Oxford: Oxford University Press.

Feinstein, J.S., Adolphs, R., Damasio, A., & Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current Biology*, 21(1), 34–38.

Feinstein, J. S., Adolphs, R., & Tranel, D. (2016). A tale of survival from the world of Patient SM. *Living Without an Amygdala*, Guilford, New York, 1–38.

Feinstein, J.S., Buzza, C., Hurlemann, R., Follmer, R. L., Dahdaleh, N. S., Coryell, W. H., ... & Wemmie, J.A. (2013). Fear and panic in humans with bilateral amygdala damage. *Nature Neuroscience*, 16(3), 270–272.

Frijda, N. H. (1986). The emotions. Cambridge University Press. Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, 14(2), 251.

Gray, J.A. (1981). A critique of Eysenck's theory of personality. In H.J. Eysenck (Ed.). *A model for personality* (pp. 246–276). Berlin and Heidelberg: Springer.

Grootswagers, T., Kennedy, B. L., Most, S. B., & Carlson, T. A. (2021). Neural signatures of dynamic emotion constructs in the human brain. *Neuropsychologia.*

Hofer, P. A. (1973). Urbach-Wiethe disease (lipoglycoproteinosis; lipoid proteinosis; hyalinosis cutis et mucosae). A review. Acta dermato-venereologica. *Supplementum*, 53, 1–52.

Hubel, D. H., & Wiesel, T. N. (1979). Brain mechanisms of vision. *Scientific American*, 241(3), 150–163.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, 99(3), 561–565.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on psychological science*, 2(3), 260–280.

Kirby, L. A., & Robinson, J. L. (2017). Affective mapping: An activation likelihood estimation (ALE) meta-analysis. *Brain and cognition*, 118, 137–148.

Krack, P., Kumar, R., Ardouin, C., Dowsey, P.L., McVicker, J.M., Benabid, A.-L., & Pollak, P. (2001). Mirthful laughter induced by subthalamic nucleus stimulation. *Movement Disorders*, 16(5), 867–875.

Kragel P. A., & LaBar KS. (2014). Advancing emotion theory with multi- variate pattern classification. *Emotion Review*, 6(2), 160–174.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84(3), 394–421.

LeDoux J. E. (2012). Rethinking the emotional brain. *Neuron*, 73(4), 653– 676.

Lin, D., Boyle, M. P., Dollar, P., Lee, H., Lein, E. S., Perona, P., & Anderson, D. J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. *Nature*, 470(7333), 221–226.

Lindquist, K., Wager, T., Kober, H., Bliss-Moreau, E., & Barrett, L. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, 35, 121–143.

MacLean, P., (1952). Some psychiatric implications of physiological studies on fronto-temporal portion of limbic system (visceral brain). *Electroencephalography & Clinical Neurophysiology*, 4, 407–418.

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel pattern analysis for FMRI data: a review. *Computational and mathematical methods in medicine*, 2012.

Menon, V. (2015). Large-scale functional brain organization. *Brain mapping: An encyclopedic reference*, 2, 449–59.

Mineka, S., & Öhman, A. (2002). Born to fear: non-associative vs associative factors in the etiology of phobias. *Behaviour research and therapy*, 40(2), 173–184.

Motzkin, J. C., Philippi, C. L., Wolf, R. C., Baskaya, M. K., & Koenigs, M. (2015). Ventromedial prefrontal cortex is critical for the regulation of amygdala activity in humans. *Biological psychiatry*, 77(3), 276–284.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, affective, & behavioral neuroscience*, 3(3), 207–233.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, 108, 483–522.

Okun, M.S., Bowers, D., Springer, U., Shapira, N.A., Malone, D., Rezai, A.R., et al. (2004). What's in a "smile"? Intra-operative observations of contralateral smiles induced by deep brain stimulation. *Neurocase*, 10, 271–279.

Panksepp, J. (1994). The basics of basic emotion. In P. Ekman & R.J. Davidson (Eds.). *The nature of emotion: Fundamental questions* (pp. 20–24). New York: Oxford University Press.

Panksepp, J. (1998). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *Journal of consciousness studies*, 5(5-6), 566–582.

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of life reviews*, 11(3), 400–435.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16(2), 331–348.

Pitzalis, S., Strappini, F., Bultrini, A., & Di Russo, F. (2018). Detailed spatiotemporal brain mapping of chromatic vision combining high−resolution VEP with fMRI and retinotopy. *Human brain mapping*, 39(7), 2868–2886.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.

Ryou, J. W., Cho, S. Y., & Kim, H. T. (2001). Lesions of the entorhinal cortex impair acquisition of hippocampal-dependent trace conditioning. *Neurobiology of learning and memory*, 75(2), 121–127.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., ... & Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral cortext*, 26(6), 2563–2573.

Schneirla, T. C. (1959). An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal.

Seeley, W. W. (2019). The salience network: a neural system for perceiving and responding to homeostatic demands. *Journal of Neuroscience*, 39(50), 9878–9882.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 32(4), 811–830.

Skinner, B. F. (1984). The evolution of behavior. *Journal of the experimental analysis of behavior*, 41(2), 217.

Tettamanti, M., Rognoni, E., Cafiero, R., Costa, T., Galati, D., & Perani, D. (2012). Distinct pathways of neural coupling for different basic emotions. *Neuroimage*, 59(2), 1804–1817.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and sociobiology*, 11(4-5), 375–424.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of cognitive neuroscience*, 22(12), 2864–2885.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS ComputBiol*, 11(4), e1004066.